ОБРАБОТКА СИГНАЛОВ И ИЗОБРАЖЕНИЙ

УДК 62-50

ОПТИМИЗАЦИЯ ПАРАМЕТРОВ В ПРОСТРАНСТВЕ БАЗИСНЫХ ФУНКЦИЙ ДЛЯ ЗАДАЧИ СЛЕЖЕНИЯ[1]

Гнеушев А.Н.
Учреждение Российской академии наук
Вычислительный центр им. А.А. Дородницына РАН, г. Москва

Поступила в редакцию 14.06.2009, после переработки 01.07.2009.

В статье развивается подход построения текстурно-геометрической модели заданного класса деформируемых объектов на основе разложения их изображения по системе базисных функций. Усовершенствован метод оптимизационного получения параметров базисных функций для заданного множества изображений. Представлено обобщение метода оценивания параметров деформации изображения объекта, заключающееся в оптимизации невязки модели в собственной системе координат в базисном пространстве. Строится текстурно-геометрическая аффинная модель лица человека с использованием семейства базисных функций Габора (Gabor Wavelet Network) для задачи слежения за лицом по видеоизображениям в режиме реального времени.

In article the approach of texture-geometrical model construction of the set class of deformable objects on the basis of decomposition images on system of basic functions develops. The method of optimizing parameters reception of basic functions for the set of images is improved. Generalization of the estimation parameters method of object deformation image is presented, consisting in optimization models discrepancy in own system of co-ordinates in basic space. The texture-geometrical affine model of the person using Gabor Wavelet Network for a problem of tracking the person under video images in a real-time is under construction.

Ключевые слова: оптимизация, базисные функции Габора, задача слежения.
Keywords: optimization, Gabor Wavelet Network, tracing problem.

Введение

Построению текстурно-геометрической модели изображения объекта со сложной структурой, такой, как лицо человека, уделяется большое внимание в связи с возросшим интересом к проблеме распознавания личности по видеоизображениям. Для применения систем распознавания в реальных условиях необходимо

[1]Работа выполнена при финансовой поддрежке РФФИ (проект № 05-07-90096).

решение задач автоматического обнаружения объекта, отдельных его элементов, масштаба и ракурса, что позволяет выделить характерные области изображения для распознавания.

Формализация частной задачи анализа изображения сводится к построению модели рассматриваемого класса объектов при заданных условиях наблюдения. Специфика объектов определяет методы обработки изображения для выделения оптимально характеризующих их признаков. Подходы к описанию текстурных объектов условно можно разделить на интегральные и локальные (иерархические). При интегральном подходе рассматриваются характеристики изображения всего объекта [1], [2], при локальном подходе изображение разбивается на подобласти, которые в совокупности характеризуют объект [3]–[6]. Локальные признаки достаточно точны и инвариантны к нелинейным изменениям геометрии объекта, однако чаще всего характеризуют неуникальные элементы. Интегральные признаки устойчивы к шумам, но относительно неточны как из-за возможного нелинейного изменения геометрии изображения, так и вследствие того, что интегральная модель является обобщающей, представляет объекты некоторого класса.

Будем рассматривать изображения естественных объектов (таких как лица людей), имеющих сложную и неоднородную структуру (текстуру) областей. Изображения таких объектов целесообразно описывать на основе аппроксимации целых областей некоторым семейством функций [7], выбранных таким образом, чтобы наиболее эффективно представлять структуру (текстуру) изображения. Эффективность в данном случае понимается как использование наименьшего количества необходимых функций для представления изображения с наилучшим качеством. Критерий качества, как правило, выбирается на основе меры близости представления к исходному изображению объекта и служит характеристикой степени точности (грубости) построенной модели.

Широкое практическое развитие получили подходы построения текстурных моделей на основе разложения изображения по различным семействам вейвлет-функций, в частности по семейству функций Габора [1]–[6], [8]. Фиксированный набор фильтров Габора привлекается в работах [3], [8], [9]. Отклики фильтров представляют яркостные признаки текстуры, ориентация фильтров и их положение характеризуют ее геометрические свойства. Развивается так же подход построения функционального базиса с помощью оптимизационных методов, что позволяет уменьшить количество базисных функций по сравнению с применением фиксированных фильтров [1], [2], [4]–[6].

В работе, опираясь на результаты исследований [1], [5], [6], рассматривается построение текстурно-геометрической модели заданного класса изображений объектов путем их представления в пространстве базисных вейвлет-функций. Используется общий функциональный базис, единый для всех изображений объектов заданного класса и представляющий каждое такое изображение как совокупность весов разложения (образ). Процесс локализации целевого объекта состоит в минимизации невязки проекции деформированного анализируемого изображения на базисное пространство и образа модели по параметрам преобразования. Оптимизация производится в модельной системе координат меньшего масштаба, чем исходное изображение методом доверительных областей [6], [10].

1. Представление изображений заданного класса объектов

Рассмотрим задачу аппроксимации областей изображения некоторым семейством функций, выбранных таким образом, чтобы наиболее эффективно представлять текстуру изображения целевых объектов. Пусть $f(\mathbf{x})$ — функция изображения, где $\mathbf{x} = (x, y)^{\mathrm{T}}$. Будем ее рассматривать как функцию из пространства $L^2(\Re^2)$ и искать ее среднеквадратичное представление в виде конечного ряда

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i}(\mathbf{x}), \quad \mathbf{n}_i \in \mathbf{\Omega_n}, \tag{1}$$

где $\psi_{\mathbf{n}_i}(\mathbf{x})$ — базисные функции в пространстве $L^2(\Re^2)$ с вектором параметров $\mathbf{n}_i$, $\mathbf{\Omega_n}$ — множество допустимых векторов $\mathbf{n}_i$, w_i — веса разложения по базисным функциям, N — число базисных функций, $\hat{f}$ — проекция изображения $f(\mathbf{x})$ на пространство базисных функций $\{\psi_{\mathbf{n}_i}\}_{i=1,\ldots,N}$. Выражение (1) можно записать в векторной форме для $\mathbf{\Psi} = (\psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_N})^{\mathrm{T}}$, $\mathbf{w} = (w_1, \ldots, w_N)^{\mathrm{T}}$: $\hat{f}(\mathbf{x}) = \mathbf{\Psi}^{\mathrm{T}}\mathbf{w}$.

Параметры базисных функций $\mathbf{n}_i$ находятся путем минимизации среднеквадратичного отклонения изображения целевого объекта и разложения (1)

$$\varepsilon^2(\mathbf{\Psi}) = \min_{\mathbf{n}_i \in \mathbf{\Omega_n},\, w_i} \left\| f - \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i} \right\|_2^2, \tag{2}$$

где ε — среднеквадратическая ошибка (точность) представления (1). При заданном векторе $\mathbf{\Psi}$ вектор весов $\mathbf{w}$ находится путем проецирования изображения на систему базисных функций $\{\psi_{\mathbf{n}_i}\}_{i=1,\ldots,N}$:

$$\mathbf{w} = \mathbf{G}_{\mathbf{\Psi}^{-1}}(\mathbf{\Psi}(\mathbf{x}), f(\mathbf{x})), \tag{3}$$

где $\mathbf{G_\Psi} = \left(\left\langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \right\rangle\right)$ — матрица Грама, составленная из скалярных произведений базисных функций, $(\mathbf{\Psi}(\mathbf{x}), f(\mathbf{x})) = \left(\left\langle \psi_{\mathbf{n}_i}, f \right\rangle, \ldots, \left\langle \psi_{\mathbf{n}_N}, f \right\rangle\right)^{\mathrm{T}}$ — вектор, каждый элемент которого есть скалярное произведение изображения с соответствующей базисной функцией $\psi_{\mathbf{n}_i}$, $\langle \cdot, \cdot \rangle$ — операция скалярного произведения относительно $\mathbf{x}$.

Конкретный выбор семейства базисных функций $\{\psi_{\mathbf{n}_i}\}_{i=1,\ldots,N}$ определяется тем, какие особенности (признаки) необходимо извлечь из анализируемого изображения. Вследствие того, что в целом образ объекта на изображении локален и состоит из элементов, имеющих статистически-локальную связь, в области анализа изображений особое место занимает класс вейвлет-функций, хорошо локализованных как в пространственной, так и в частотной области. Семейство вейвлет-функций образует базис в $L^2(\Re^2)$ и любая функция $f \in L^2(\Re^2)$ с любой степенью точности и единственным образом может быть представлена с помощью интегрального вейвлет-преобразования в этом базисе [11]. Семейство двумерных вейвлет-функций порождаются из единственной материнской вейвлетной функции путем определения ее параметров $\mathbf{n}_i$, таких как сдвиг, масштаб, поворот, которые задают ее «степени свободы» и позволяют «настраивать» каждую базисную функцию на определенную характерную область изображения.

В (1) используется только конечное число базисных функций. Таким образом, одному разложению (1) (проекции) соответствует несколько исходных функций изображения. В этом случае эти изображения будут близки по норме в $L^2(\Re^2)$ и, следовательно, ряд (1) будет представлять некоторый класс изображений.

Определение 1. Классом изображения объектов C_ε относительно нормы $\|\cdot\|$ определим такое множество изображений $\{f_k\}$, что для $\forall\, i,\, j\ \exists\, \varepsilon$: $\|f_i - f_j\| \le \varepsilon$.

Отсюда видно, что при $\varepsilon_1 < \varepsilon$ класс C_{ε_1} является подклассом C_ε: $C_{\varepsilon_1} \subset C_\varepsilon$. Таким образом, разложение (1) представляет класс изображений C_ε относительно нормы $\|\cdot\|_2$ с ε, определяемой по формуле (2). Из выражения (2) следует, что параметры $\mathbf{n}_i$, а значит, и базис $\mathbf{\Psi}$ зависят от конкретного изображения объекта, т.е. класс C_ε порождается конкретным изображением.

Утверждение. Множество классов $\{C_{\varepsilon_k}\}$, которые порождаются каждым изображением f_k в соответствии с выражением (2), является подклассом класса $C_{\varepsilon_{\mathbf{\Psi}}}$, для которого

$$\varepsilon_{\mathbf{\Psi}}^2 = \min_{\mathbf{n}_i \in \mathbf{\Omega}_\mathbf{n}, w_i^k} \left(\sum_{k=1}^{K} \left\| f_k - \sum_{i=1}^{N} w_i^k \psi_{\mathbf{n}_i} \right\|_2^2 \right), \tag{4}$$

где f_k — изображение с номером k из рассматриваемого множества, K — общее количество изображений во множестве, $\mathbf{w}^k$ — вектор весов разложения изображения f_k по базису. Действительно, выражение (4) можно записать как $\varepsilon_{\mathbf{\Psi}}^2 = \sum_{k=1}^{K} \varepsilon_k^2(\mathbf{\Psi})$, где $\varepsilon_k(\mathbf{\Psi})$ — точность представления f_k в едином базисе $\mathbf{\Psi}$. Пусть $\mathbf{\Psi}_k$ — оптимальный базис для изображения f_k, полученный из выражения (2), тогда $\varepsilon_k(\mathbf{\Psi}_k) \le \varepsilon_k(\mathbf{\Psi}) \le \varepsilon_{\mathbf{\Psi}}$.

Для описания геометрически деформируемых изображений объектов расширим определение класса.

Определение 2. Пусть на множестве рассматриваемых изображений определены система координат M_{cs} и ее непрерывное взаимнооднозначное преобразование $T_\mathbf{a}(\mathbf{x})$: $\Re^2 \to \Re^2$ с вектором параметров $\mathbf{a}$. Классом изображения объектов $C_{\varepsilon,T}$ относительно нормы $\|\cdot\|$ и преобразования $T_\mathbf{a}$ определим такое множество изображений $\{f_k(\mathbf{x})\}$, что для $\forall i,\, j\ \exists\ \mathbf{a}(i,j)$, $\varepsilon(\mathbf{a})$: $\|f_i(\mathbf{x}) - f_j(T_\mathbf{a}(\mathbf{x}))\| \le \varepsilon$.

Из определения класса $C_{\varepsilon,T}$ следует, что для $\forall f_k \in C_{\varepsilon,T}\ \exists \mathbf{a}(k)$: $\{f_k(T_{\mathbf{a}(k)}(\mathbf{x}))\} \subseteq C_{\varepsilon_\mathbf{\Psi}}$, т.е. множество изображений класса $C_{\varepsilon_\mathbf{\Psi},T}$ можно отобразить в класс $C_{\varepsilon_\mathbf{\Psi}}$, где значение $\varepsilon_\mathbf{\Psi}$ определяется из выражения (4). Представление (1) изображения объекта из класса $C_{\varepsilon_\mathbf{\Psi},T}$ тем лучше, чем меньше значение $\varepsilon_\mathbf{\Psi}$ при заданном количестве базисных функций N и изображений K. Таким образом, эффективность описания изображения зависит от выбора семейства базисных функций, выбора набора их параметров и адекватностью преобразования $T_\mathbf{a}$, моделирующего деформации изображения объектов рассматриваемого класса.

2. Текстурно-геометрическая модель

Определим текстурно-геометрическую модель для изображений объектов из класса $C_{\varepsilon,T}$ относительно взаимно однозначного и непрерывно дифференцированного преобразования $T_\mathbf{a}$: $\Re^2 \to \Re^2$. Будем предполагать, что функции преобразования $T_\mathbf{a}(\mathbf{x})$ ограничены и равномерно непрерывны со всеми своими частными производными, а якобиан прямого и обратного преобразования отличен от нуля на области, ограниченной изображением объекта. Введем модельную систему координат M_{cs}, в которой параметризуем геометрические особенности объекта на изображении некоторым множеством характерных точек $\mathbf{P}_M = \{\mathbf{x}_m\}$. Таким образом, текстурно-геометрическую модель изображения объекта зададим множеством

$$M_O = \left\{\mathbf{\Psi},\, \mathbf{w}^k,\, \mathbf{P}_M,\, T_{\mathbf{a}}\right\}_{k=0,\dots,K}, \tag{5}$$

где $\mathbf{\Psi}$ — вектор базисных функций, представляющий с вектором весов $\mathbf{w}^k$ образ k; K — количество образов объекта, содержащихся в модели; $T_{\mathbf{a}}$ — преобразование системы координат, моделирующее деформации изображения объектов класса $C_{\varepsilon,T}$.

Модель M_O строится с помощью следующей обучающей процедуры. На исходных изображениях объектов $\{f_k^O\}_{k=1,\dots,K}$ эксперт задает множество характерных точек $\mathbf{P}_O^k = \{\mathbf{x}_o^k\}$. Далее формируется обучающая выборка $\{f_k\}_{k=1,\dots,K}$, путем преобразования: $\{f_k^O(\mathbf{x})\} \to \{f_k(T_{\mathbf{a}(k)}(\mathbf{x}))\}$, где $T_{\mathbf{a}(k)}$: $\mathbf{P}_M \to \mathbf{P}_O^k$. Обучение модели (5) состоит в нахождении параметров $\mathbf{n}_i$ базисных функций $\psi_{\mathbf{n}_i}$ и весов разложения $\mathbf{w}^k = (w_1^k, \dots, w_N^k)^{\mathrm{T}}$, оптимально аппроксимирующих обучающую выборку изображений $\{f_k\}_{k=1,\dots,K}$, с помощью выражения (4). Задача (4) имеет большую размерность. Для упрощения разобьем ее на несколько подзадач и будем искать базисное разложение итеративно [6]. Оптимизация производится для каждой новой функции $\psi_{\mathbf{n}_t}$ и веса w_t^k на остатке, который определяется разностью обучающего множества изображений и уже найденным базисным рядом с $\mathbf{\Psi}_{t-1}$ и $\mathbf{w}_{t-1}^k$ на предыдущей итерации

$$\varepsilon_t^2 = \min_{\mathbf{n}_t \in \mathbf{\Omega}_{\mathbf{n}},\, w_t^k} \sum_{k=1}^{K} \left\|(f_k - \mathbf{\Psi}_{t-1}^{\mathrm{T}} \mathbf{w}_{t-1}^k) - w_t^k \psi_{\mathbf{n}_t}\right\|_2^2, \quad \mathbf{w}_t^k = \mathbf{G}_{\mathbf{\Psi}_t}^{-1} \left\langle f_k, \mathbf{\Psi}_t \right\rangle, \tag{6}$$

где $\mathbf{\Psi}_t = (\psi_{\mathbf{n}_1}, \dots, \psi_{\mathbf{n}_t})^{\mathrm{T}}$, $\mathbf{w}_t^k = (w_1^k, \dots, w_t^k)^{\mathrm{T}}$, $t \le N$. Можно показать, что полученная таким образом система функций $\{\psi_{\mathbf{n}_i}\}_{i=1,\dots,N}$ линейно независима [1]. В работе минимизация (6) осуществляется на основе метода доверительных областей [6], [10].

3. Оценка геометрических параметров объекта

Пусть имеется обученная модель (5) и некоторое изображение $f(\mathbf{x})$, содержащее целевой объект. Оценка параметров объекта заключается в получении множества характерных точек $\mathbf{P}_O = \{\mathbf{x}_o\}$ в системе координат анализируемого изображения $f(\mathbf{x})$. Для этого находятся такие параметры $\mathbf{a}$ преобразования $T_{\mathbf{a}}$, при котором $\mathbf{P}_M \to \mathbf{P}_O$. Оценка вектора $\mathbf{a}$ производится путем минимизации невязки проекции трансформированного изображения $f(T_{\mathbf{a}}(\mathbf{x}))$ на систему базисных функций и весовых векторов $\mathbf{w}^k$ в пространстве и системе координат M_{cs} модели. Минимизация невязки может быть записана с учетом (3) в виде

$$\min_{\substack{\mathbf{a} \in \mathbf{\Omega}_{\mathbf{a}} \\ k=1,\dots,K}} \left\|\mathbf{w}^k - \mathbf{G}_{\mathbf{\Psi}}^{-1}(\mathbf{\Psi}(\mathbf{x}), f(T_{\mathbf{a}}(\mathbf{x})))\right\|_{\mathbf{\Psi}}^2, \tag{7}$$

где $\mathbf{\Omega}_{\mathbf{a}}$ — множество допустимых векторов $\mathbf{a}$, задающее ограничения на преобразование, $\mathbf{w}^k$ — весовой вектор образа k, представленный в модели. Оптимизация (7) осуществляется по норме $\|\mathbf{w}-\mathbf{v}\|_{\mathbf{\Psi}}$ (где $\mathbf{w}$ и $\mathbf{v}$ — вектора в пространстве базисных функций), которую можно определить как [1]

$$\|\mathbf{w}-\mathbf{v}\|^2_{\mathbf{\Psi}} = \left\| \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i} - \sum_{j=1}^{N} v_j \psi_{\mathbf{n}_j} \right\|^2_2 = (\mathbf{w}-\mathbf{v})^{\mathrm{T}} \mathbf{G}_{\mathbf{\Psi}} (\mathbf{w}-\mathbf{v}).$$

Таким образом, находится такое преобразование $\mathrm{T}^{-1}_{\mathbf{a}}$ системы координат изображения объекта, чтобы его образ попал в класс обучающей выборки $C_{\varepsilon_{\mathbf{\Psi}}}$. В [5], [6] используется норма $\|\mathbf{w}-\mathbf{v}\|^2_{\bar{\mathbf{\Psi}}} = \|\mathbf{w}/\|\mathbf{w}\|_{\mathbf{\Psi}} - \mathbf{v}\|\mathbf{v}\|_{\mathbf{\Psi}}\|^2_{\mathbf{\Psi}}$, $\|\mathbf{w}\|_{\mathbf{\Psi}} \neq 0$, $\|\mathbf{v}\|_{\mathbf{\Psi}} \neq 0$, пропорциональная нормированному коэффициенту корреляции весовых векторов и не учитывающая локальные мультипликативные изменения яркости изображения.

В работе оптимизация (7) осуществляется с помощью итеративного метода доверительных областей [6]. Для инициализации процесса (7) задается начальный вектор $\mathbf{a}$ на основе приближенной оценки характерных точек $\mathbf{P}_O$ на изображении объекта, например, с помощью грубой его локализации некоторым алгоритмом выделения области интереса. Метод доверительных областей состоит в решении следующей квадратичной подзадачи на каждой итерации i [10]

$$\min_{d_i \in \mathbf{\Omega}} \mathbf{g}_i^{\mathrm{T}} \mathbf{d}_i + 1/2 \mathbf{d}_i^{\mathrm{T}} \mathbf{A}_i \mathbf{d}_i, \quad \|\mathbf{D}\mathbf{d}_i\|_2 \leq \Delta_i, \tag{8}$$

где $\mathbf{g}_i = -2\mathbf{J}^{\mathrm{T}}(\mathbf{a}_i)\mathbf{G}_{\mathbf{\Psi}}(\mathbf{w}^k - \mathbf{v}(\mathbf{a}_i))$ — градиент функции невязки для (7); $\mathbf{v}(\mathbf{a}_i) = \mathbf{G}^{-1}_{\mathbf{\Psi}}(\mathbf{\Psi}(\mathbf{x}), f(T_{\mathbf{a}_i}(\mathbf{x})))$ — образ трансформированного изображения в базисе; $\mathbf{A}_i = 2\mathbf{J}^{\mathrm{T}}(\mathbf{a}_i)\mathbf{G}_{\mathbf{\Psi}}\mathbf{J}(\mathbf{a}_i)$ — симметричная матрица, аппроксимирующая матрицу Гессе; $\mathbf{J}(\mathbf{a})$ — матрица Якоби $\|\partial\mathbf{v}(\mathbf{a})/\partial\mathbf{a}\|$; $\mathbf{d}_i$ — решение подзадачи, приращение к текущему вектору параметров: $\mathbf{a}_{i+1} = \mathbf{a}_i + \mathbf{d}_i$; $\mathbf{D}$- диагональная матрица, задающая ограничение в эллипсоиде (доверительную эллипсоидную область); Δ_i — размер доверительной области.

Параметр k (номер образа $\mathbf{w}^k$ в модели) не входит в вектор параметров $\mathbf{a}_i$ для задачи (8). Для его оценки на каждой итерации i выбирается такой образ k в модели M_O (5), который наиболее близок к проекции изображения $\mathbf{v}(\mathbf{a}_i)$ на текущей итерации: $\min_{k=1,\dots,K} \|\mathbf{w}^k - \mathbf{v}(\mathbf{a}_i)\|^2_{\mathbf{\Psi}}$.

Возможность градиентной оптимизации в системе координат модели M_{cs} достигается благодаря ограничениям, накладываемым на преобразование $T_{\mathbf{a}}$ для модели (5). Матрица Якоби $\mathbf{J}(\mathbf{a})$ находится аналитически в системе координат исходного изображения с помощью замены переменных в выражении

$$\mathbf{v}(\mathbf{a}) = \mathbf{G}^{-1}_{\mathbf{\Psi}}(\mathbf{\Psi}(\mathbf{x}), f(\mathrm{T}_{\mathbf{a}}(\mathbf{x}))) = \mathbf{G}^{-1}_{\mathbf{\Psi}}\left(\left|\mathbf{J}\left[\mathrm{T}^{-1}_{\mathbf{a}}(\mathbf{x})\right]\right| \mathbf{\Psi}(\mathrm{T}^{-1}_{\mathbf{a}}(\mathbf{x})), f(\mathbf{x})\right),$$

где $\left|\mathbf{J}\left[\mathrm{T}^{-1}_{\mathbf{a}}(\mathbf{x})\right]\right|$ — модуль якобиана преобразования $\mathrm{T}^{-1}_{\mathbf{a}}$. Далее, полученная матрица Якоби с помощью обратной замены переводится в систему координат модели M_{cs}, где она зависит от неизменных в процессе итераций базисных функций $\psi_{\mathbf{n}}(\mathbf{x})$ и их частных производных, которые, таким образом, можно рассчитать заранее, еще на этапе обучения модели, что существенно увеличивает вычислительную экономичность минимизации (7). Заметим, что положение, размеры и ориентация целевого объекта на исходном изображении f не влияют на скорость оптимизации, так как все операции производятся в системе координат модели M_{cs}.

4. Текстурно-геометрическая модель изображения лица и слежение за ним

Рассмотрим построение модели изображения лица на основе изложенного подхода. Одним из ключевых моментов к эффективному описанию изображения объекта является выбор семейства базисных функций. Широкое практическое развитие для анализа изображений получило семейство вейвлет-функций Габора [1]–[6], [8]. Они оптимально локализованы как в пространственной, так и в частотной области [8], [11] и вследствие этого обладают хорошей частотно-пространственной селективностью. Хотя, в общем случае, функции Габора не являются вейвлетами, однако можно выделить такой подкласс функций этого семейства, который удовлетворяет условию «допустимости вейвлетов» [8], [11]. Вследствие того, что базисные вейвлет-функции локальны и имеют нулевое среднее, при свертке в пределах носителя каждой базисной функции не учитываются постоянные составляющие изображения. Данные составляющие можно рассматривать как кусочную аппроксимацию неравномерной освещенности, к которой образ в таком базисе устойчив. В [9] показано, что фильтр на основе функций этого семейства является также хорошим инструментом для выделения текстурных признаков изображения. Таким образом, в качестве базисных функций для (1) выберем, как и в [1], [6], семейство нечетных функций Габора в виде

$$\begin{gathered}\psi_{\mathbf{n}_i}(x,y)=\sqrt{\frac{2}{(1-e^{-\pi^2})\pi^3}}\exp\left[-\frac{(S_i)^2}{2\pi^2}(x_{ri}^2+(\gamma_i)^2y_{ri}^2)\right]\sin(S_ix_{ri}),\\ x_{ri}=(x-X_i)\cos\theta_i+(y-Y_i)\sin\theta_i,\\ y_{ri}=-(x-X_i)\sin\theta_i+(y-Y_i)\cos\theta_i\end{gathered}\tag{9}$$

с параметрами $\mathbf{n}_i=(X_i,Y_i,S_i,\gamma_i,\theta_i)$ и ограничениями:

$$\mathbf{\Omega_n}=\{\mathbf{n}\colon 0.1<S<1,\ 0.5<\gamma<2\}\,.\tag{10}$$

Ограничения (10) определены с использованием биологической модели зрительной клетки [8]. Функция (9) принадлежит пространству $L^2(\Re^2)$ и удовлетворяет условию «допустимости вейвлетов» [1], [11].

Будем рассматривать изображения лиц как изображения из класса $C_{\varepsilon,T}$ относительно аффинного преобразования $T_{\mathbf{a}}(\mathbf{x})=\mathbf{SR}(\mathbf{x}-\mathbf{c})$, где $\mathbf{S}=\begin{pmatrix}s_x & s_{xy}\\ 0 & \gamma s_x\end{pmatrix}$, $\mathbf{R}=\begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}$, $\mathbf{c}=(c_x,c_y)^{\mathrm{T}}$, $\mathbf{x}=(x,y)^{\mathrm{T}}$, $\mathrm{s}_x>0$, $\gamma>0$ с вектором параметров $\mathbf{a}=(c_x,c_y,s_x,\gamma,s_{xy},\varphi)^{\mathrm{T}}$.

Зададим систему координат модели M_{cs} и сформируем обучающую выборку изображений путем аффинного отображения исходных изображений в систему координат M_{cs} в соответствии с экспертной расстановкой характерных точек лица, таких как глаза и рот. Оптимизация (6) производится с помощью градиентного метода доверительных областей [6]. Для инициализации параметров $\mathbf{n}_i$ функции Габора распределяются по области изображения объекта «слоями»: сначала область лица равномерно покрывается низкочастотными функциями (с малым значениями S), затем с более высокой частотой и т.д. Компоненты векторов $\mathbf{w}^k$ задаются единицами. Для уменьшения краевых эффектов в базисном пространстве границы области обучения на изображениях сглаживаются гауссовским фильтром.

Вследствие того, что аффинное преобразование определяет перемещение, поворот и масштабирование системы координат, построенная модель (5) позволяет решать задачу слежения за лицом по видеоизображениям. Для первого изображения в последовательности задается оценка вектора $\mathbf{a}$, полученного с помощью формирования аффинного преобразования трех опорных точек из системы координат M_{cs} в систему координат изображения, на котором они грубо могут быть найдены помощью, например, детектора лица [12]. На каждом последующем изображении для минимизации (7) в виде начального приближения задается вектор $\mathbf{a}$, оцененный на предыдущем изображении. Оценку положения характерных точек лица в системе координат текущего изображения получаем путем преобразования: $\mathbf{P}_O = T_{\mathbf{a}}(\mathbf{P}_M)$.

Устойчивость и скорость сходимости (7) определяется начальным выбором вектора $\mathbf{a}$. На практике получены следующие оценки точности задания $\mathbf{a}$ для успешной глобальной сходимости: сдвиг от истинного положения объекта составляет 27% размера целевого объекта; размер по отношению к истинному устанавливается таким, чтобы обеспечить 3-кратное различие в масштабе в большую или меньшую сторону; поворот от истинной ориентации объекта составляет $\pm 40°$. В проведенных экспериментах использовалась модель изображения лица размером 50×50 точек, содержащая 70 базисных функций и обученная по 200 изображениям лиц. На компьютере с процессором класса Pentium IV 2.8 GHz слежение за лицом при около фронтальных умеренных движениях производилось стабильно в режиме реального времени при 30 кадрах в секунду.

Список литературы

[1] Feris R., Krueger V., Cesar Jr.R. Efficient real-time face tracking in wavelet subspace Proc. of IEEE ICCV Workshop on Recog., Anal. and Tracking of Faces and Gestures in Real-time Systems. Vancouver, 2001.

[2] Hu C., Feris R., Turk M. Active wavelet networks for face alignment British Machine Vision Conf. Norwich, 2003.

[3] Wiskott L., Fellous J. M., Krueger N. et al. Face recognition by elastic bunch graph matching. IEEE Trans. Patt. Anal. and Mach. Intel. 1997. V. 7, № 19.

[4] Feris R., Gemmell J., Toyama K. et al. Hierarchical wavelet networks for facial feature localization Proc. of IEEE ICCV Workshop on Recog., Anal. and Tracking of Faces and Gestures in Real-time Systems. Vancouver, 2001.

[5] Гнеушев А.Н. Локализация элементов лица путем оптимизации разложения изображения по базисным функциям Габора. Теоретические и прикладные задачи нелинейного анализа. М.: ВЦ РАН, 2005.

[6] Гнеушев А.Н. Построение и оптимизация текстурно-геометрической модели изображения лица в пространстве базисных функций Габора. Известия Академии Наук. Теория и системы управления. 2007. № 3.

[7] Ту Дж., Гонсалес Р. Принципы распознавания образов. М.: Мир, 1978.

[8] Lee T.S. Image representation using 2D Gabor wavelets IEEE Trans. Patt. Anal. and Mach. Intel. 1996. V. 18, № 10.

[9] Manjunath B.S., Chellappa R. A unified approach to boundary perception: edges, textures, and illusory contours. IEEE Trans. neural networks. 1993. V. 1, № 4.

[10] Byrd R., Schnabel R.B., Shultz G.A. A trust region algorithm for nonlinearly constrained optimization SIAM J. Numer. Anal. 1987. № 24.

[11] Чуи Ч. Введение в вейвлеты. М.: Мир, 2001.

[12] Viola P., Jones M. Robust real-time face detection International J. Computer Vision. 2004. V. 57, № 2.